INAF radio data archive

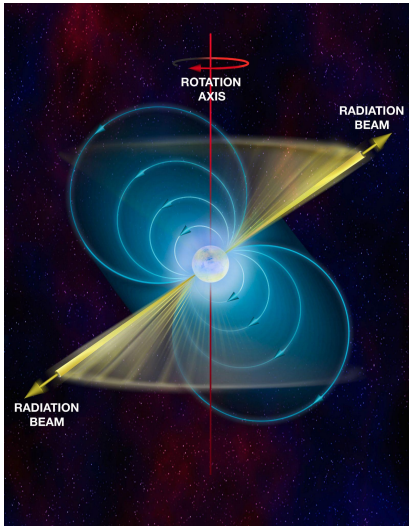
ongoing activities & future plans about pulsar/FRB data

Coordinators: *A. Zanichelli, C. Knapic (IA2)*

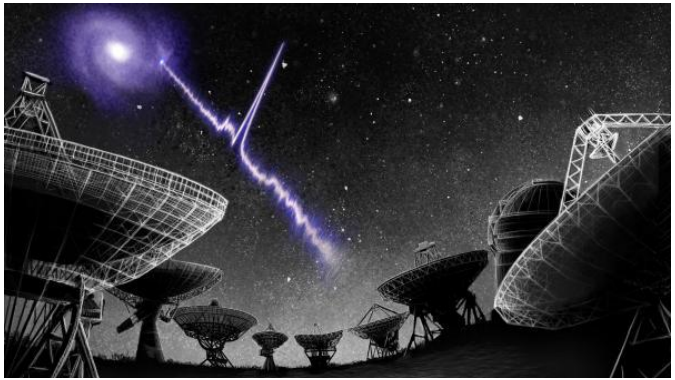
People: *V. Galluzzi, S. Zorba, M. Molinaro, A. A. R. Fara, M. Burgay*

Summary

1. Pulsar/transients observations and data formats
2. Data archiving workflow
3. Data model
4. Use stories examples
5. Summary of current activities and future plans



Credit: B. Saxton, NRAO/AUI/NSF



Credit: Astron/Jive/Evn, Daniëlle Futselaar (artsource.nl)

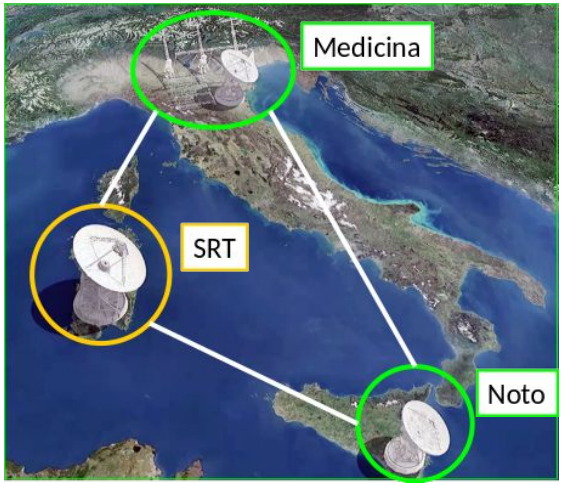
Pulsar/transients observations

BACK-ENDS

Name available bandwidths (MHz)	Type
TP 250, 680, 1200 (C and K bands only), 2000 (C and K bands only)	Analog total power
XARCOS Narrow band spectrometer; up to four (for single-feeds) simultaneous bandwidths: 0.5, 2.0, 7.8, 62.5	Spectro-polarimeter
SARDARA(e) 420, 1000 (L-band only, no f-track), 1500	Spectro-polarimeter
DFB3 1024	Correlator for pulsars
ROACH1(f) 128 (8 x 16 MHz)	Baseband recorder
DBBC 512	VLBI

RECEIVERS

RF band (GHz)	Type	Offered for
P 0.30-0.36(c) **L** 1.3-1.8	Dual frequency coaxial feed, cryo-cooled	VLBI, single-dish
C-high 5.7-7.7	Single-feed, cryo-cooled	VLBI, single-dish
K 18-26.5	7-feed, cryo-cooled	VLBI, single-dish

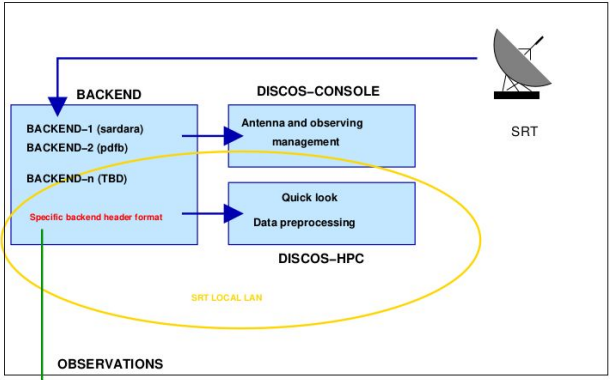


- 2 regular call for proposals per year (deadlines in April and October)
- ~ 2.1 TB raw data collected during the period 2018-2020 with PDFB3 (~ 470 files published so far)
- SARDARA spectrometer: effective data rate 20-30 GByte/hour
- on average, 150 MB to 10 GB per data file in folding mode acquisition, 3 to few hundreds GBs in search mode
- proprietary period of 1 yr (except from long programmes and/or particular requests from PIs)
- For long-term preservation, 2.5 PB tape library by IBM (equipped with LTO 8 cartridges)
- More information available at https://www.radiotelescopes.inaf.it/

Pulsar/transients data formats

BACK-ENDS

Name available bandwidths (MHz)	Type
TP 250, 680, 1200 (C and K bands only), 2000 (C and K bands only)	Analog total power
XARCOS Narrow band spectrometer; up to four (for single-feeds) simultaneous bandwidths: 0.5, 2.0, 7.8, 62.5	Spectro- polarimeter
SARDARA(e) 420, 1000 (L-band only, no f-track), 1500	Spectro- polarimeter
DFB3 1024	Correlator for pulsars
ROACH1(f) 128 (8 x 16 MHz)	Baseband recorder
DBBC 512	VLBI



baseband data
(.dada)

PSRFITS | *.sf* *.cf* | *.rf*

dspsr
psrchive
other sw lib

splice_dada
digifil (dspsr)
other sw

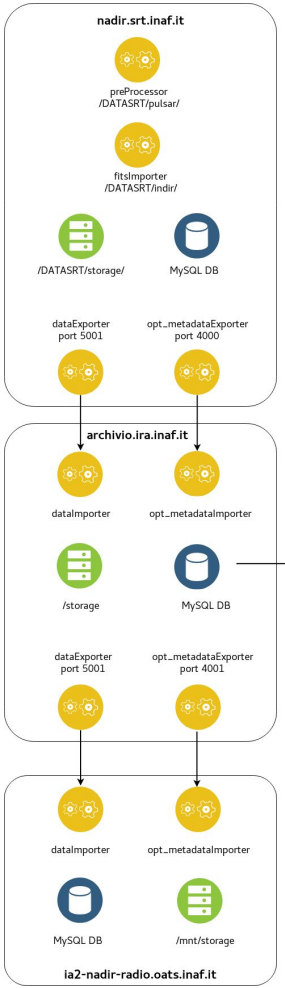
FILTERBANK
(.fil)

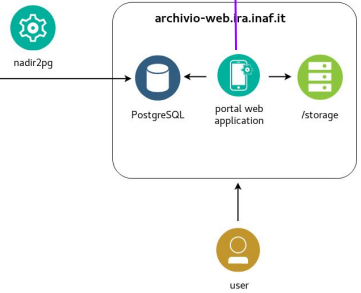
online

soon online

offline

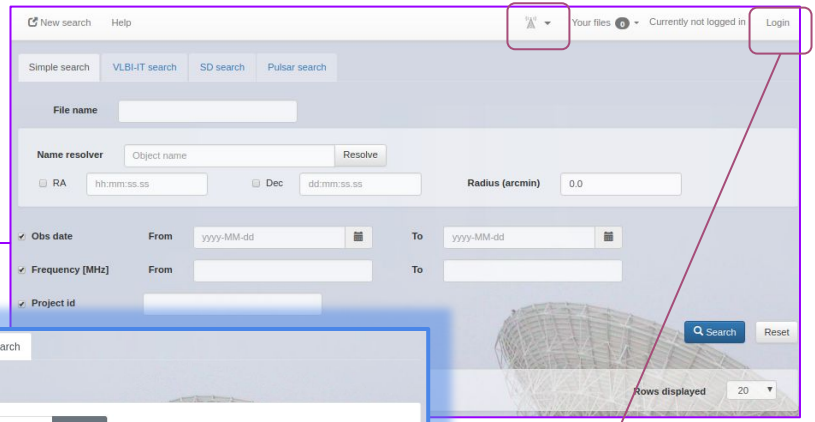
The archival system

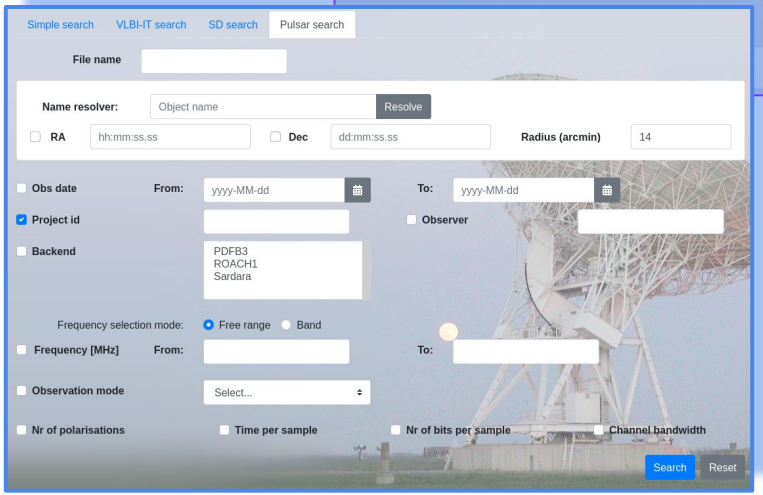


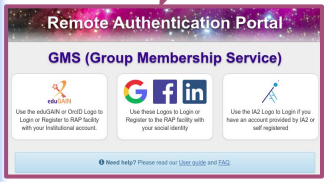


The **archival system** is based on the ***New Archiving Distributed InfrastructuRe*** **(NADIR)**, explicitly designed to be **flexible** in order to cope with **evolving data models, formats, publication policies, versions and metadata contents, keeping consistencies** among different sites.

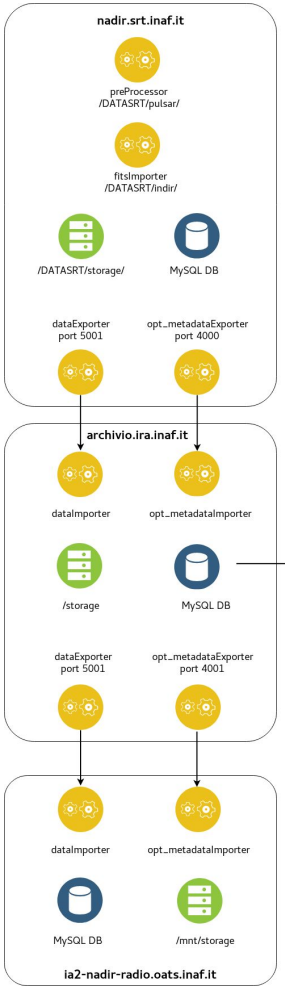
SAMP broadcast



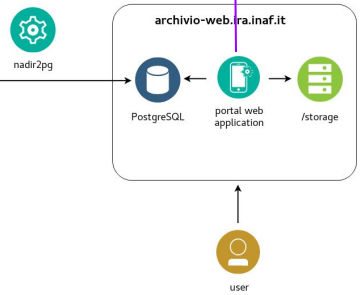




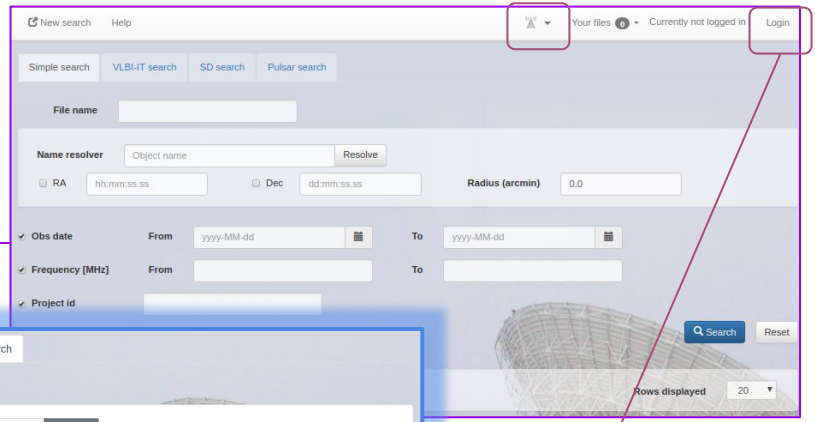
The archival system

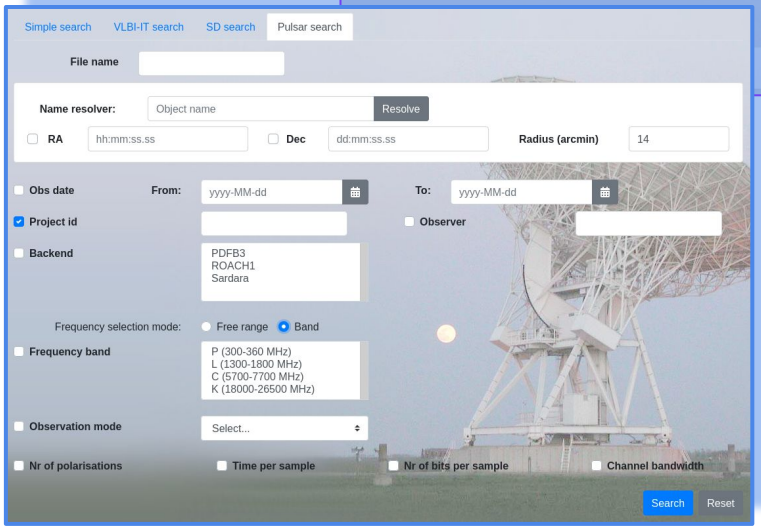


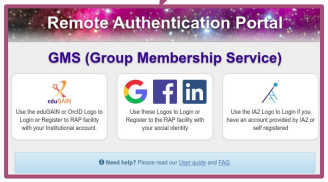
The **archival system** is based on the ***New Archiving Distributed InfrastructuRe*** **(NADIR)**, explicitly designed to be **flexible** in order to cope with **evolving data models, formats, publication policies, versions and metadata contents, keeping consistencies** among different sites.



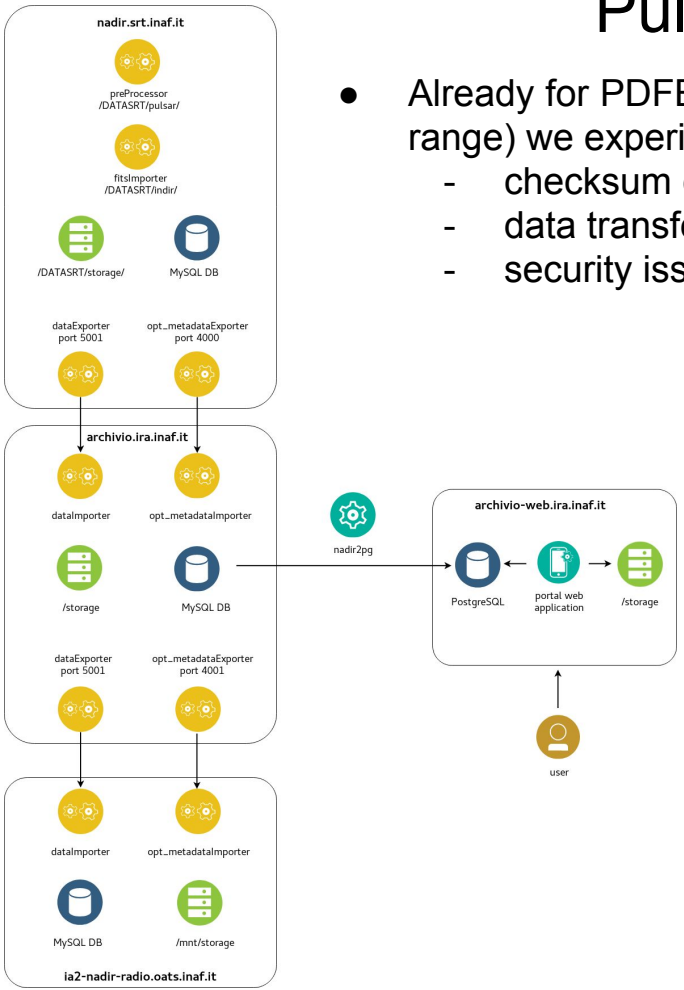
SAMP broadcast







Pulsar ingestion issues



- Already for PDFB3 backend observations (PSRFITS with sizes in the 10 MB - 10 GB range) we experienced CPU overload. This might depend on:
  - checksum calculations of tar.gz archives for GB-class FITS files;
  - data transfers from SRT (Cagliari) to IRA (Bologna)
  - security issues (e.g. Log4j library)

- As we will also have FILTERBANK (both pre-processing and ingestion) from ROACH-1 and ROACH-2 (Sardara) backends (up to few hundreds GBs in file size), we are planning tests in order to better investigate and, then, to try to fix such problems.

Pulsar data model

- The metadata of an observation are all written in the header of the primary HDU of a FITS file (in case of FILTERBANK, we produce an accompanying FITS file containing only a primary header).

```
SIMPLE  =                    T / file does conform to FITS standard
BITPIX  =                    8 / number of bits per data pixel
NAXIS   =                    0 / number of data axes
EXTEND  =                    T / FITS dataset may contain extensions
COMMENT   FITS (Flexible Image Transport System) format is defined in 'Astronomy
[MORE COMMENT LINES]
HDRVER  = '4.0               '   / Header version
FITSTYPE= 'PSRFITS           '   / FITS definition for pulsar data files
DATE    = '2018-12-07T17:59:10' / file creation date (YYYY-MM-DDThh:mm:ss UT)
OBSERVER= 'Marta Burgay      '   / Observer name(s)
PROJID  = '34-18   '             / Project name
TELESCOP= 'SRT     '             / Telescope name
ANT_X   =              4865206.61169 / [m] Antenna ITRF X-coordinate (D)
ANT_Y   =               791927.345226 / [m] Antenna ITRF Y-coordinate (D)
ANT_Z   =              4035152.25842 / [m] Antenna ITRF Z-coordinate (D)
FRONTEND= 'Our Frontend'     / Rx and feed ID
NRCVR   =              2 / Number of receiver polarisation channels
FD_POLN = 'CIRC    '         / LIN or CIRC
FD_HAND =                    1 / +/- 1. +1 is LIN:A=X,B=Y, CIRC:A=L,B=R (I)
FD_SANG =             45. / [deg] FA of E vect for equal sig in A&B (E)
FD_XYPH =             23. / [deg] Phase of A^* B for injected cal (E)
BACKEND = 'PDFB3   '             / Backend ID
BECONFIG= 'pdfb4_1024_512_2048' / Backend configuration file name
BE_PHASE=                    -1 / 0/+1/-1 BE cross-phase:0 unknown,+/-1 std/rev
BE_DCC  =                    0 / 0/1 BE downconversion conjugation corrected
BE_DELAY=                    0. / [s] Backend propn delay from digitiser input
TCYCLE  =             10. / [s] On-line cycle time (D)
OBS_MODE= 'PSR     '         / (PSR, CAL, SEARCH)
DATE-OBS= '2018-12-07T17:59:20' / Date of observation (YYYY-MM-DDThh:mm:ss UTC)
OBSFREQ =             6206. / [MHz] Centre frequency for observation
OBSBW   =             512. / [MHz] Bandwidth for observation
OBSNCHAN=                    2048 / Number of frequency channels (original)
CHAN_DM =                    0. / [cm-3 pc] DM used for on-line dedispersion
SRC_NAME= 'B0355+54'         / Source or scan ID
COORD_MD= 'J2000   '         / Coordinate mode (J2000, GAL, ECLIP, etc.)
EQUINOX = '2000.000'         / Equinox of coords (e.g. 2000.0)
```

```
RA      = '03:58:54.717'     / Right ascension (hh:mm:ss.ssss)
DEC     = '+54:13:13.727'    / Declination (-dd:mm:ss.sss)
BMAJ    =  0.0547856912665163 / [deg] Beam major axis length
BMIN    =  0.0547856912665163 / [deg] Beam minor axis length
BPA     =             0. / [deg] Beam position angle
STT_CRD1= '03:58:54.717'     / Start coord 1 (hh:mm:ss.sss or ddd.ddd)
STT_CRD2= '+54:13:13.727'    / Start coord 2 (-dd:mm:ss.sss or -dd.ddd)
TRK_MODE= 'TRACK   '         / Track mode (TRACK, SCANGC, SCANLAT)
STP_CRD1= '03:58:54.717'     / Stop coord 1 (hh:mm:ss.sss or ddd.ddd)
STP_CRD2= '+54:13:13.727'    / Stop coord 2 (-dd:mm:ss.sss or -dd.ddd)
SCANLEN =            139.810 / [s] Requested scan length (E)
FD_MODE = 'FA      '         / Feed track mode - FA, CPA, SPA, TPA
FA_REQ  =            0.2 / [deg] Feed/Posn angle requested (E)
CAL_MODE= 'SYNC    '         / Cal mode (OFF, SYNC, EXT1, EXT2)
CAL_FREQ=                    0. / [Hz] Cal modulation frequency (E)
CAL_DCYC=                    0. / Cal duty cycle (E)
CAL_PHS =            0. / Cal phase (wrt start time) (E)
STT_IMJD=            58459 / Start MJD (UTC days) (J - long integer)
STT_SMJD=            64760 / [s] Start time (sec past UTC 00h) (J)
STT_OFFS=            0.1312937734375 / [s] Start time offset (D)
STT_LST =            86399.8683467561 / [s] Start LST (D)
STT_DATE= '2018-12-07'
STT_TIME= '17:59:20.000'
HIERARCH OBSDATAFORMAT = 'PSRFITS ' / Data format of the observation
HIERARCH SUBINT.NPOL =     4 / Nr of polarisations
HIERARCH SUBINT.TBIN = 0.00015272180983952 / [s] Time per bin or sample
HIERARCH SUBINT.NBITS =    1 / Nr of bits/datum (SEARCH mode 'X' data, else 1)
HIERARCH SUBINT.CHAN_BW = 0.25 / [MHz] Channel/sub-band width
COMMENT This file has been verified and finalized via the
COMMENT pre-processor script checkfits_radio.sh (version 3.7)
[MORE COMMENT LINES]
```

Pulsar data model

- The metadata of an observation are all written in the header of the primary HDU of a FITS file (in case of FILTERBANK, we produce an accompanying FITS file containing only a primary header).
- Then, the metadata are mapped onto a flat table (datamodel_pulsar) in a MySQL db (metadata_pulsar).

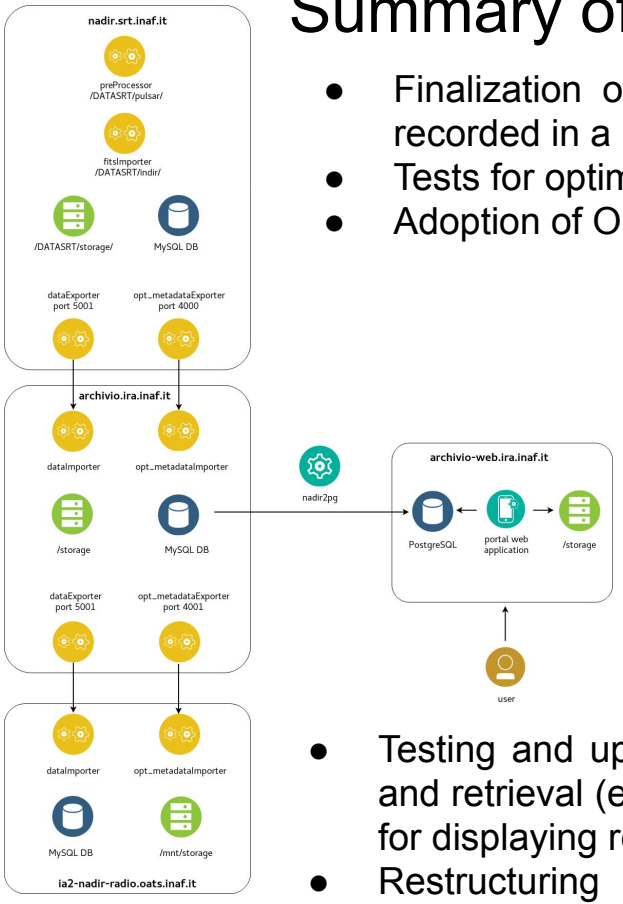
id	column_name	column_type	fits_key_hdu	fits_key_pri	fits_key_sec	
1	TELESCOP	varchar	0	TELESCOP	TELESCOP	Telescope name
2	DATE_OBS	varchar	0	DATE-OBS	DATE-OBS	Date of observation (YYYY-MM-DDThh:mm:ss UTC)
3	SRC_NAME	varchar	0	SRC_NAME	SRC_NAME	Source or scan ID
4	OBSERVER	varchar	0	OBSERVER	OBSERVER	Observer name(s)
5	OBS_MODE	varchar	0	OBS_MODE	OBS_MODE	PSR, CAL, SEARCH)
6	BACKEND	varchar	0	BACKEND	BACKEND	Backend ID
7	RA_C	varchar	0	RA	RA	Right ascension (hh:mm:ss.ssss)
8	DEC_C	varchar	0	DEC	DEC	Declination (-dd:mm:ss.sss)
9	EQUINOX	double	0	EQUINOX	EQUINOX	Equinox of coords (e.g. 2000.0)
10	PROJID	varchar	0	PROJID	PROJID	Project name
11	OBSFREQ	double	0	OBSFREQ	OBSFREQ	[MHz] Centre frequency for observation
12	**OBSBW**	**double**	**0**	**OBSBW**	**OBSBW**	**[MHz] Bandwidth for observation**
13	**SCANLEN**	**double**	**0**	**SCANLEN**	**SCANLEN**	**[s] Requested scan length (E) (N.B.: diff. from MBFITS)**
14	**NPOL**	**int**	**0**	**SUBINT.NPOL**	**SUBINT.NPOL**	**Nr of polarisations**
15	**TBIN**	**double**	**0**	**SUBINT.TBIN**	**SUBINT.TBIN**	**[s] Time per bin or sample**
16	**NBITS**	**int**	**0**	**SUBINT.NBITS**	**SUBINT.NBITS**	**Nr of bits/datum (SEARCH mode 'X' data, else 1)**
17	**CHAN_BW**	**double**	**0**	**SUBINT.CHAN_BW**	**SUBINT.CHAN_BW**	**[MHz] Channel/sub-band width**
18	**OBSDATAFORMAT**	**vachar**	**0**	**OBSDATAFORMAT**	**OBSDATAFORMAT**	**Data format of the observation**

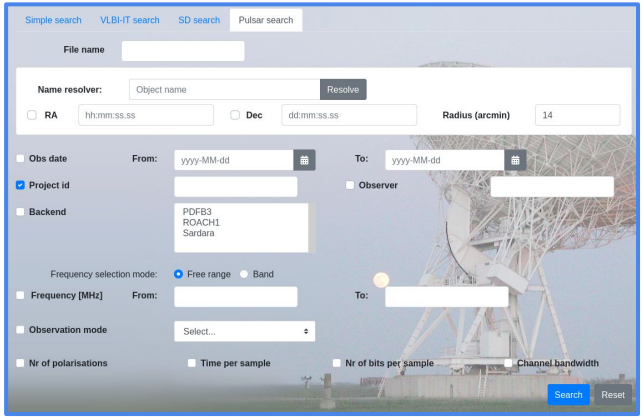
Pulsar/transients science user stories examples

- Reprocessing archival data may lead to discovery of new pulsars (e.g. high DM and/or short period ones) simply missed by previous studies, due to some inefficiency of previous algorithms or, more trivially, large number of candidates to scrutinize in case of pulsar surveys. [Mickalinger+2012, Pan+2016]. The same applies for FRBs (Lorimer+2007,Keane+2019).
- A radio scientist, e.g. in performing an observation at a given frequency of an already known pulsar, typically wants to complement his/her data with other radio frequencies (possibly full-Stokes observations).
- A pulsar candidate identified from optical of even gamma observations may be confirmed by existing radio observations (e.g. it is a known pulsar, Nieder+2020). Also, a null result from radio archives may bolster the science case for new proposals.
- Binary pulsars or a pulsar with a stellar black hole companion are useful to constraint gravity in its strong regime (Kramer+2021). Such studies typically require long term monitoring programmes.
- Again, from a more theoretical perspective, searching pulsar data by directly using physical parameters (DM, RM, P, Pdot, B, binary system) may help in selecting a particular class of pulsar for dedicated statistical studies (e.g. to constraint magnetic fields and radiative processes in that particular class, understand evolutive processes).
- The number of detected FRBs is rapidly growing (from a few tens in 2018 to a few hundredths in 2021). Being events located at cosmological distances (z up to ~ 3, DM > 1000 pc/cm^3), these naturally offers as cosmological probes (Ze-Wei Zhao+2020,Wucknits+2021).

Summary of ongoing activities & future plans

- Finalization of the script for generating a FITS summary file for an observation recorded in a FILTERBANK file
- Tests for optimization of ingestion procedures
- Adoption of ObsCore DM (and evaluating CAOM)





File name	Policy	Project id	Frequency [MHz]	Backend	Nr of polarisations	Time per sa
20181211-215206-34-18-FRB121102.sf.gz	FREE	34-18	6462	PDFB3	1	1.25E-4
20181211-222524-34-18-FRB121102.sf.gz	FREE	34-18	6462	PDFB3	1	1.25E-4
20181211-233236-37-18-B0355+54.rf.gz	FREE	37-18	18512	PDFB3	4	6.10891598:
20181211-234208-37-18-B0355+54.sf.gz	FREE	37-18	18512	PDFB3	1	1.0E-4
20181211-235318-37-18-J1023+0038.sf.gz	FREE	37-18	18512	PDFB3	1	1.0E-4
20181212-094054-37-18-B1937+21.rf.gz	FREE	37-18	1804	PDFB3	4	6.08547743-
20181215-120203-37-18-B0355+54.rf.gz	FREE	37-18	18512	PDFB3	4	3.05447157(
20181215-120641-37-18-B0355+54.rf.gz	FREE	37-18	18512	PDFB3	4	3.05447154(
20181215-121657-37-18-B0355+54.rf.gz	FREE	37-18	18512	PDFB3	4	3.05447150.
20181215-122149-37-18-B0355+54.rf.gz	FREE	37-18	18512	PDFB3	4	3.05447148.
20181215-123331-37-18-B0355+54.rf.gz	FREE	37-18	18512	PDFB3	4	3.05447143-
20181215-124429-37-18-B0355+54.rf.gz	FREE	37-18	18512	PDFB3	4	3.05447139.
20181215-132531-45-18-B1933+16.rf.gz	FREE	45-18	320	PDFB3	4	7.00706025.
20181215-133353-45-18-B1937+21.rf.gz	FREE	45-18	320	PDFB3	4	3.04273213.
20181215-133607-45-18-B1937+21.rf.gz	FREE	45-18	320	PDFB3	4	3.04273216

- Testing and upgrades of web interfaces in order to enhance data discovery, access and retrieval (e.g. additional query fields and columns for results, more readable layout for displaying results)
- Restructuring the currently available TAP service (by complementing available information and, more generally, simplifying the database schema)
- Future plans to deliver calibrated as well as more advanced data products